Measuring Inefficiency in Public Utilities: Does the Distribution Matter?

Martín Rossi and Iván Canay

Working Paper N° 12
ISBN 987-519-086-1
(May 2001)

CEER
Centro de Estudios Económicos de la Regulación
Departamento de Economía y Finanzas, Universidad Argentina de la Empresa
Lima 717, 1° piso
C1053AAO Buenos Aires, Argentina
Telephone: 54-11-43797693
Fax: 54-11-43797588
E-mail: ceer@uade.edu.ar
http://www.uade.edu.ar/economia/ceer.

(Please see the back pages of this paper for a listing of current working paper titles and subscription ordering information).

The Centro de Estudios de Economía de la Regulación (CEER), is an organization engaged in the analysis of the public regulation of utilities. The CEER is supported by the World Bank, the Regulatory Agencies of the Public Utilities of Water and Electricity of Argentina, and the Universidad Argentina de la Empresa (Buenos Aires), where the CEER works.

CEER Working Paper Series
Measuring Inefficiency in Public Utilities: Does the Distribution Matter?
Martín Rossi and Iván Canay
Working Paper N° 12
(May, 2001)
JEL N°: L5, L9

Abstract: In this paper we describe two of the most used distributions for the inefficiency term in a stochastic frontier: the Half Normal and the Exponential distributions. We then show that both distributions affect the skewness of the composite error term in different ways, which make the Exponential distribution more inclined to identifying a larger number of efficient firms. In order to check these results we perform MOLS and ML, using four databases already used in previous work. We find that the mean efficiency is sensitive to the assumed distribution. In all cases we corroborate that the Exponential distribution identifies a larger number of efficient firms than the Half Normal, as the theory predicted. However, the rankings of firms by efficiency scores were not affected by the choice of distribution.

Affiliation of the author: Martín Rossi (Alumni Oxford University, Formerly CEER/UADE), Iván Canay, Centro de Estudios Económicos de la Regulación (CEER), Universidad Argentina de la Empresa (UADE), Buenos Aires, Argentina.

CEER
Centro de Estudios Económicos de la Regulación
Departamento de Economía y Finanzas, Universidad Argentina de la Empresa
Lima 717, 1° piso
C1053AAO Buenos Aires, Argentina
Telephone: 54-11-43797693
Fax: 54-11-43797588
E-mail: ceer@uade.edu.ar
http://www.uade.edu.ar/economia.ceer.

Measuring Inefficiency in Public Utilities: Does the Distribution Matter?

Martín A. ROSSI
Centro de Estudios Económicos de la Regulación - UADE, Argentina.

Iván A. CANAY
Centro de Estudios Económicos de la Regulación - UADE, Argentina.

Abstract

In this paper we describe two of the most used distributions for the inefficiency term in a stochastic frontier: the Half Normal and the Exponential distributions. We then show that both distributions affect the skewness of the composite error term in different ways, which make the Exponential distribution more inclined to identifying a larger number of efficient firms. In order to check these results we perform MOLS and ML, using four databases already used in previous work. We find that the mean efficiency is sensitive to the assumed distribution. In all cases we corroborate that the Exponential distribution identifies a larger number of efficient firms than the Half Normal, as the theory predicted. However, the rankings of firms by efficiency scores were not affected by the choice of distribution.

JEL classification: L5, L9.

I. Introduction

Since the mid 1990s, with the benefit of a deeper understanding of the potential benefits of yardstick competition between regional monopolies (Schleifer, 1985), practitioners and academics specializing in regulatory issues have increasingly become interested in developing standardized performance indicators for monopolies in the infrastructure sector.

These indicators can be used as inputs in a price cap regulatory regime with an RPI-X rule, in the measurement of the X factor. Among these indicators, efficiency frontiers have become predominant during the last years. These kinds of frontiers can be estimated with parametric and non-parametric techniques, and the distance of the observed practice to the frontier can be considered deterministic or stochastic.

In the deterministic approach, all the firms share the same frontier and the inefficiency is a residual concept given by the discrepancy between the individual firm performance and the estimated frontier. Thus, this approach completely ignores the possibility of a single firm performance being affected not only by inefficiencies in the management of its resources but also by factors absolutely beyond its control.

With the works of Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977) the so-called stochastic frontiers made their appearance. These are based on the idea that deviations from the frontier could be partially out of the control of the analyzed firm, leaving room for random noise. The stochastic approach adds the problem of the decomposition, into inefficiency and noise, of the error term. In order to perform this decomposition it is necessary to assume some distribution for both components of the composed error term.

In the last years, the literature has focused almost exclusively on stochastic frontiers. However, the following questions still remain unanswered: does the assumption made for the distribution of the inefficiency term matter in empirical work? Is the decision made by regulators in setting the X factor affected by this assumption? In this work we will try to answer these questions by addressing the theoretical and practical implications of the use of different distributions for the inefficiency component of the error term.

The paper outline is as follows. In section II we describe two of the most commonly used distributions for the inefficiency term: the Half Normal (H-N) and the Exponential distributions, and their use in stochastic frontier estimation. We then show that the Exponential distribution has a tendency to identify a larger number of efficient firms than the H-N distribution. In section III we perform estimations of two production functions and two cost functions for three different sectors, seeking empirical support for the conclusions of sections II. Finally, in section IV we make our conclusions.

II. The distributions

The cost frontier specification is given by[1]:

$$\ln C_i = \alpha + X\beta + \underbrace{v_i + u_i}_{\varepsilon_i}$$

where X is a matrix of explanatory variables, $\varepsilon_i = v_i + u_i$ is the composed error term, v_i is an unrestricted variable and u_i is the inefficiency term, which, in a cost frontier, is non-negative.[2]

The u_i component cannot be directly observed; therefore it has to be inferred from the composed error term. In order to perform this decomposition and establish which part of the composed term corresponds to random noise and which part to inefficiency, it will be necessary to assume some distribution for both components. The noise term, v_i, is less problematic, because there is a consensus that this variable is independently and identically distributed with a normal distribution $N(0,\sigma_v^2)$. On the other hand, several functional forms have been proposed for the inefficiency term: Half-Normal (Aigner, Lovell and Schmidt, 1977), Truncated Normal (Stevenson, 1980), Gamma (Green, 1990) and Exponential (Meeusen and van den Broeck, 1977). The assumption choice implies a trade-off between flexibility and simplicity. The Half-Normal (H-N) and the Exponential distributions are particular cases of the Truncated Normal and the Gamma distributions, respectively, and are easier to use since they are distributions with only one parameter. This simplicity has as a counterpart the loss of the greater flexibility provided by the Truncated Normal and the Gamma distributions. However, authors like Ritter and Simar (1997) highlight the difficulties associated with the estimation of the two parameters of these distributions, and recommend the use of relatively simpler distributions.

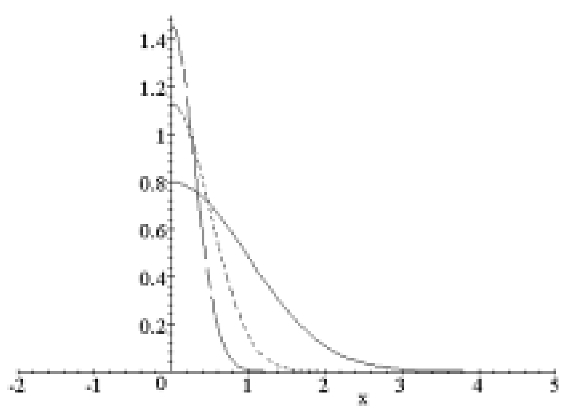
---

[1] Deterministic frontiers are particular cases of this formulation where $v_i = 0$

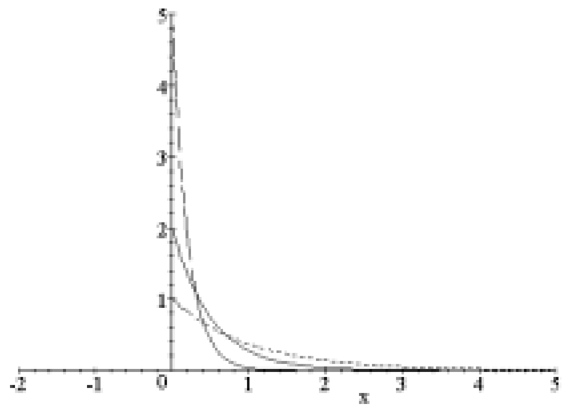
[2] In a production frontier u_i is non-positive.

In this paper we will analyze both one-parameter distributions. The following graphics show the Half-Normal[3] $N^{+}(0,\sigma_u^2)$ and the Exponential[4] $Ex(\theta)$ distribution for selected values of the parameters:

Figure 1



Half-Normal distribution for different values of σ_u



Exponential distribution for different values of θ

The graphics show that both distributions have most of their probabilistic mass near zero, and this mass increases when the variance decreases.[5] This property has an economic implication: the majority of the firms under study are almost efficient. Nevertheless, we have to keep in mind that there is no theoretical reason for inefficiency to behave always in such a way, and that it could be distributed otherwise.

Given that v_i and u_i are assumed independent, the density function of ε_i is asymmetrically distributed with mean and variance:

$$E(\varepsilon_i^{HN}) = E(u_i) = \left(\frac{2}{\pi}\right)^{1/2} \sigma_u \qquad V(\varepsilon_i^{HN}) = \frac{\pi - 2}{\pi}\sigma_u^2 + \sigma_{v(HN)}^2$$

---

[3] The density function is $f(u) = \frac{2}{\sqrt{2\pi}\sigma_u}\exp\left(-\frac{u^2}{2\sigma_u^2}\right)$

[4] The density function is $f(u) = \theta e^{-\theta u}$

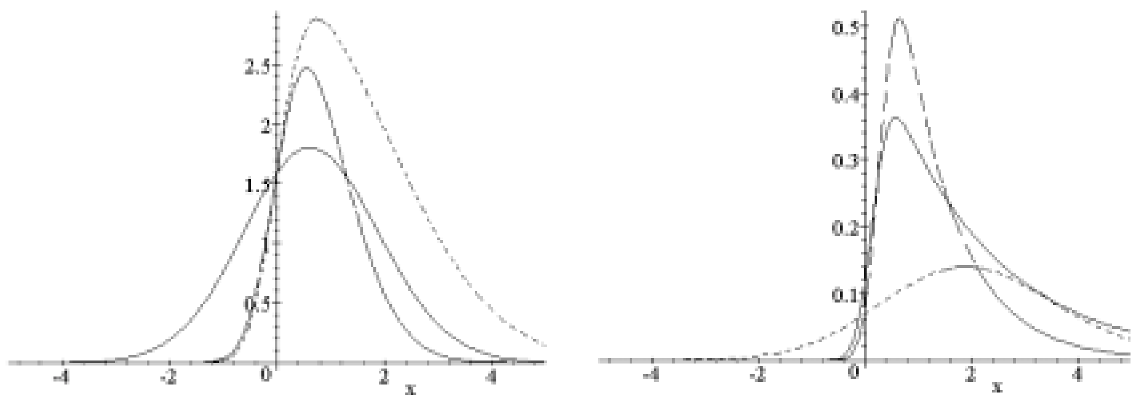
[5] The variance of the Exponential is $\frac{1}{\theta^2}$

for the Half-Normal case, and

$$E(\varepsilon_i^{Exp}) = E(u_i) = \frac{1}{\theta} \qquad V(\varepsilon_i^{Exp}) = \frac{1}{\theta^2} + \sigma^2_{v(Exp)}$$

for the Exponential case. It is not difficult to notice that the skewness of the composed error term must be positive in both cases.[6] If the skewness of the estimated error is negative, this could be showing that the data is inconsistent with the selected functional form (Waldman, 1982). This diagnostic is independent of the assumption made for the distribution of the inefficiency component.

Figure 2



Distribution of ε_i^{HN} for different values of σ_u and σ_v. Distribution of ε_i^{Exp} for different values of θ and σ_v.

Note that if the assumptions made about v_i and u_i are correct, the shape of the estimated error must be similar to those in figure 2. If the shape of the estimated ε_i looks like a H-N, this would be implying that the estimated frontier has no noise and thus it would be similar to the deterministic case. The contrary occurs if the estimated ε_i looks like a Normal distribution, because in this case the estimated frontier is a typical OLS estimation where all the firms are 100% efficient (all the distance between the observations and the frontier is attributed to random noise).

[6] The third moment is $\mu_3 = -\left(\frac{2}{\pi}\right)^{1/2}\left(1-\frac{4}{\pi}\right)\sigma_u^3$ for the H-N and $\mu_3 = \frac{2}{\theta^3}$ for the Exponential.

Stochastic frontiers must be estimated in two parts. In the first part we have to obtain consistent estimates of the technological parameters and a consistent estimate of the parameter of the assumed distribution. At this stage we can perform ML (Maximum Likelihood) and obtain all the estimates at once, or we can use MOLS (Modified Ordinary Least Squares) and make the two necessary steps to obtain consistent estimates of the intercept parameter. Finally, in the second part, the composed error term must be decomposed in order to obtain estimates of the inefficiency for each firm.

As already mentioned, MOLS[7] estimation requires two steps. The first step is independent of the assumption made for the inefficiency term and it consists basically in an OLS estimation of the cost function. Thus, we obtain consistent and unbiased estimates for the slope parameters and consistent but biased estimates for the constant term. In the second step it is necessary to make the distributional assumption for u_i and obtain estimates of $E(u_i)$ by means of the moments of the OLS residuals. Then, the biased OLS intercept is corrected using the estimated expected value, $E(u_i)$. It is important to notice that the frontier estimated with this procedure is simply the average function, shifted upwards or downward, implying that the technological parameters of the frontier are the same of the average function. This is one of the aspects that make MOLS different from ML. The latter incorporates a priori information on the distribution asymmetry of the error term, hence giving more weight to the efficient firms in the estimation of the slope parameters.[8]

In the H-N case, the moments of the OLS residuals can be used to obtain consistent estimates of σ_u^2 and $\sigma_{v(HN)}^2$. These parameters are used to shift the OLS intercept, transforming the average function into a frontier. In order to obtain the inefficiency scores, it is necessary to decompose the composed error term using the expression introduced by Jondrow et al. (1982):[9]

---

[7] See Lovell (1993).

[8] Olson, Schmidt and Waldman (1980) used a Montecarlo simulation to conclude that MOLS is more efficient when the sample is small (under 200 observations) and that ML is more efficient when the sample is larger.

[9] The expression introduced by Jondrow et al. has another parameterization in terms of λ instead of γ.

$$E\left[u_i/\varepsilon_i\right]=\mu_i^*+\sigma_*\left[\frac{\phi(\mu_i^*/\sigma_*)}{\Phi(\mu_i^*/\sigma_*)}\right] \text{ where } \begin{cases}\mu_i^*=\varepsilon_i^{HN}\gamma\ ;\ \sigma_*^2=\sigma_{v(HN)}^2\gamma \\ \sigma^2=\sigma_{v(HN)}^2+\sigma_u^2\ ;\ \gamma=\dfrac{\sigma_u^2}{\sigma^2}\end{cases}$$

As Olson et al. (1980) already mentioned, there are two difficulties when we use MOLS. The first one emerges when the skewness of the OLS residual has the incorrect sign, which causes a negative estimate of σ_u.[10] In such cases it is common to set $\sigma_u^2=0$ (which implies that $\gamma=0$) and consider a model with pure noise where all the firms are 100% efficient. The second difficulty appears when the variance of the OLS residual is smaller than the variance of u_i[11], thus rendering a negative estimate of $\sigma_{v(HN)}$.[12] In these cases it is common to set $\sigma_{v(HN)}^2=0$ (which implies $\gamma=1$). This result would be showing that the model is similar to a deterministic one, where all the distance to the frontier is due to inefficiency. Olson et al. (1980) concluded that the second difficulty makes the MOLS estimator inapplicable, although this conclusion is due to the parameterization used by these authors (in terms of $\lambda=\sigma_u/\sigma_{v(HN)}$).

In the Exponential distribution case, OLS residuals also can be used to obtain estimates of $\sigma_{v(Exp)}^2$ and θ. Here the OLS intercept would be shifted by the expected value of the Exponential distribution, which can be calculated once we have estimates of θ. Once more, the expression proposed by Jondrow et al. (1982) can be used for the decomposition, although the following changes must be taken into account: $\mu_i^*=\varepsilon_i^{Exp}-\sigma_{v(Exp)}^2/(1/\theta)$; $\sigma_*=\sigma_{v(Exp)}$.

The conditional mode of u_i given ε_i is another choice to infer the inefficiency component, although less recommended in the literature. Both types of decomposition arise from the

---

[10] This problem is independent of the distributional assumption and so it is also a problem in the Exponential case.

[11] It is important to note that the variance of u_i is $\frac{\pi-2}{\pi}\sigma_u^2$ not σ_u^2 .

conditional density function $f(u|\varepsilon)$ that in all cases is distributed as a Truncated Normal at zero (despite the fact that the mean and the variance of the Truncated Normal change with the distribution assumed for u_i).[13]

It is interesting to notice what the H-N and the Exponential distributions imply for both types of decomposition and observe the differences that exist between them. In the H-N case, the conditional mode for a cost frontier is:

$$M(u_i / \varepsilon_i^{HN}) = \begin{cases} \mu_i^* = \varepsilon_i^{HN}\gamma & \text{if} \quad \mu_i^* > 0 \\ 0 & \text{if} \quad \mu_i^* \leq 0 \end{cases}$$

Let’s explore the intuitive idea behind this expression. If $\varepsilon_i^{HN} < 0$, it is possible that u_i is not too large, which could be showing that the firm is efficient. On the other hand, it is possible that u_i is a large number whenever ε_i^{HN} is an increasing positive number. Therefore, in the H-N distribution case, it is the sign of ε_i^{HN} what defines if a firm is full efficient or not.

In the Exponential case, $\mu_i^* > 0$ implies $\varepsilon_i^{Exp} > \sigma_{v(Exp)}^2 /(1/\theta)$. Contrary to the former case, here it can happen that ε_i^{Exp} is a positive number and a zero of inefficiency is still assigned to a firm. It would be precipitated, however, to conclude that the Exponential distribution is more inclined to identifying a larger number of efficient firms solely on the basis of this result, because the H-N and the Exponential composed error terms are not equal. In order to understand why the Exponential distribution does have a sample mean efficiency larger than the H-N distribution, we need to explore the relationship between the variances of both distributions. This is done in the next three propositions, which are only valid for the MOLS estimator.

---

12 This problem could be a signal that the H-N distribution is not a good assumption.

13 For a detailed explanation of these topics, see Kumbhakar and Lovell (2000).

Proposition 1: In a cost frontier, the Exponential composed error term is smaller than the H-N composed error term : $\varepsilon_i^{HN} > \varepsilon_i^{EXP}$.

Given the relationship between the variance of the H-N distribution and the Exponential distribution, follows that $\frac{1}{\theta} = \kappa\sigma_u$ where κ is just the number $\kappa = \left(\frac{1}{2}\right)^{1/3}\left(\frac{2}{\pi}\right)^{1/6}\left(\frac{4}{\pi}-1\right)^{1/3}$.

This implies that: $E(u_i^{HN}) = \left(\frac{2}{\pi}\right)^{1/2}\sigma_u > E(u_i^{Exp}) = \kappa\sigma_u$ and hence the proposition 1 is always true.

Proposition 2: The ratio of the variance of noise to the total residual variance is larger in the exponential case: $\sigma^2_{v(Exp)} > \sigma^2_{v(HN)}$.

Making use of the relationship in proposition 1, follows that $\sigma^2_{v(Exp)} = \sigma^2_{v(HN)} + \left[\frac{\pi-2}{\pi} - \kappa^2\right]\sigma_u^2$. This expression implies that $\sigma^2_{v(Exp)} > \sigma^2_{v(HN)}$ because the term between brackets is positive.

Proposition 3: The Exponential distribution is more inclined to identifying a larger number of efficient firms than the H-N.

We have seen that in the exponential case $\mu_i^* > 0$ implies $\varepsilon_i^{Exp} > \sigma^2_{v(Exp)}/(1/\theta)$. If we use proposition 1 and 2 we can prove that this expression is equivalent to:

$$\varepsilon_i^{HN} > \frac{1}{\kappa}\frac{\sigma^2_{v(HN)}}{\sigma_u} + \left[\frac{(2\pi)^{1/2}\kappa - 2 + \pi(1-2\kappa^2)}{\kappa\pi}\right]\sigma_u$$

The right-hand side of the inequality is positive given that the number between brackets is positive. Thus, while the H-N could be assigning a $e^{\mu_i^*} > 1$ efficiency score to a positive

ε_i^{HN} , the Exponential distribution could be assigning a score of 1 (full efficiency) for the same ε_i^{HN} . This result is also valid for the conditional expectation.

III. Does the distribution matter?

Up to this point, nothing has indicated that one distribution is better than the other, or whether both distributions give different results. The existence of many alternatives does pose a problem, and this problem would be even worst if different assumptions gave inconsistent results. Therefore, the question remains: does the assumption made for the distribution of the inefficiency term matter in empirical work?

In order to see how the assumed distribution influences the outcomes, we will use four databases already used in previous work focused in public utilities: 1) Rossi (2001); 2) Estache and Rossi (1999); 3) Stewart (1993) and 4) CEER (2000).

Model 1

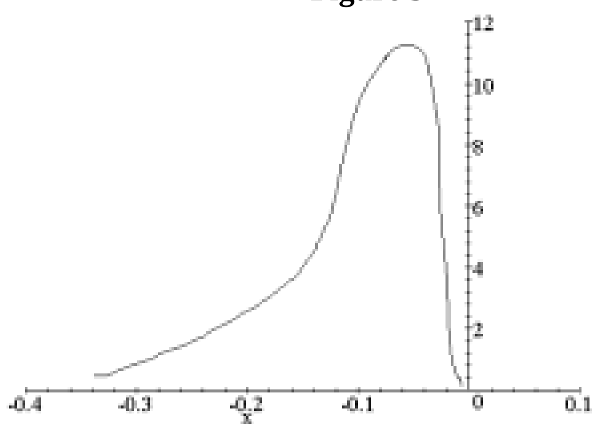
The data set used by Rossi (2001) is an unbalanced panel data of eight gas distribution firms in Argentina. The set includes information on one output, two inputs and three environmental variables. The variables are: number of customers (Custom, the output), kilometers of pipes (Kmnet, a proxy of capital input), number of employees (Empl, the labour input), concession area (Area), market structure (the ratio of residential sales to total sales, Struct) and maximum demand (Maxdem). Here we use the same variables, although we treat the model like a cross-section one. This implies that each firm observations are assumed independent over time. The MOLS and ML estimates of the production function are presented in the next table:

Model 1 (Gas - Argentina): the dependent variable is Lncustom					
Variable	**OLS**	**MOLS-HN**	**MOLS-EXP**	**ML-HN**	**ML-EXP**
Lnkmnet	0.529 (6.54)	0.529	0.529	0.664 (4.26)	0.711 (7.68)
Lnempl	0.441 (4.49)	0.441	0.441	0.362 (2.33)	0.334 (4.39)
Lnarea	-0.161 (-12.04)	-0.161	-0.161	-0.182 (-9.11)	-0.193 (-14.26)
Lnmaxdem	0.089 (1.60)	0.089	0.089	0.048 (0.47)	0.033 (0.61)
Lnstruct	-0.173 (-4.12)	-0.173	-0.173	-0.174 (-2.46)	-0.168 (-4.13)
Constant	6.361 (13.08)	6.454	6.417	6.321 (8.19)	6.308 (16.77)
Gamma	- -	0.84	- -	1	- -
% of noise	- -	34.79%	59.05%	0%	5.16%

The ratio of the estimated coefficient to the standard error is presented between parentheses.
Ln stands for natural logarithm

Note that the ML-HN estimator founds no noise in the error term[14], and for this reason the estimated frontier is similar to a deterministic one. As we mentioned before, this should be the outcome when the shape of the estimated error term looks like an H-N distribution. To check this, we plot a graphic based on the composed error term histogram.[15] The Jarque-Bera statistic gives a value of 10.45, so the null that the distribution is normal can be rejected (see Figure 3).

Figure 3



Estimated error term

[14] The proportion of noise was estimated as $\sigma_v^2/(\sigma_v^2 + \sigma_u^2\,(\pi\text{-}2)/\pi)$ for the H-N and $\sigma_v^2/(\sigma_v^2 + 1/\theta^2)$ for the Exponential distribution.

[15] Remember that the inefficiency is non-positive in a production frontier.

Model 2

Estache and Rossi (1999) estimated a cost frontier for 50 water companies in Asia and the Pacific Region. The variables included were costs (as dependent variable), wages (Salar), number of clients (Clien), daily production (Prod), number of connections (Cone), population density in the served area (Dens), percentage of water from surface sources, number of hours of water availability (Quali), metering (Meter); and a set of qualitative variables on the type of treatment used: chlorination (Dumclo), and desalination (Dumdes). In our analysis we decided to exclude all variables with a t statistic less than unity. Therefore, the model is:

Model 2 (Water- Asia and the Pacific Region): the dependent variable is Lncost					
Variable	**OLS**	**MOLS-HN**	**MOLS-EXP**	**ML-HN**	**ML-EXP**
Lnsalar	0.288 (7.18)	0.288	0.288	0.292 (7.58)	0.286 (7.99)
Lnclien	0.733 (11.15)	0.733	0.733	0.723 (11.38)	0.734 (11.31)
Lncone	0.281 (5.35)	0.281	0.281	0.303 (7.06)	0.282 (7.02)
Lndens	-0.156 (-2.49)	-0.156	-0.156	-0.148 (-1.88)	-0.153 (-1.71)
Quali	0.036 (3.55)	0.036	0.036	0.035 (3.04)	0.036 (3.15)
Meter	0.207 (1.25)	0.207	0.207	0.210 (1.50)	0.211 (1.35)
Constant	0.840 (1.37)	0.586	0.688	0.289 (0.46)	0.720 (0.95)
Gamma	- -	0.44	- -	0.78	- -
% of noise	- -	77.26%	85.72%	47.15%	95.06%

The ratio of the estimated coefficient to the standard error is presented between parentheses.
Ln stands for natural logarithm

As we can see, the ML and MOLS estimates in the table are quite close, which can be explained by the fact that the ratio of the variance of noise to the total residual variance is high (% of noise).

Model 3

Stewart (1993) performed a cost frontier analysis for 32 water companies in England. He found that the cost drivers for this sector were: volume of water sold (Sales), kilometers of pipes (Kmnet), market structure (Struct) and average pumping (Pump). Due to lack of data, we have replaced the pumping 1993 data with 1997 information. Likewise, given that four observations are missing, the sample only has 28 observations. If we take these differences into account, the estimated cost frontier is:

Model 3 (Water - England): the dependent variable is Lncost					
Variable	**OLS**	**MOLS-HN**	**MOLS-EXP**	**ML-HN**	**ML-EXP**
Lnsales	0.752 (5.21)	0.752	0.752	0.715 (4.41)	0.752 (4.69)
LnKmnet	0.442 (5.26)	0.442	0.442	0.472 (5.09)	0.440 (5.40)
Lnstruct	-0.215 (-1.98)	-0.215	-0.215	-0.209 (-1.71)	-0.214 (-1.79)
Lnpump	0.133 (1.65)	0.133	0.133	0.176 (2.13)	0.135 (1.66)
Constant	-4.314 (-10.33)	-4.417	-4.376	-4.748 (-13.1)	-4.360 (-10.7)
Gamma	- -	0.68	- -	0.95	- -
% of noise	- -	56.89%	72.93%	12.45%	81.63%

The ratio of the estimated coefficient to the standard error is presented between parentheses.
Ln stands for natural logarithm

Model 4

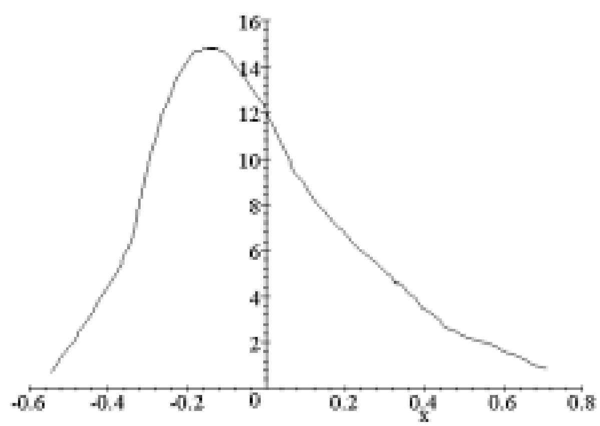
For the last estimation we will use the data used in CEER (2000). In this work the authors performed an international comparison between energy distribution firms of South America. The database is an unbalanced panel data with 99 observations for a total of 35 companies in the period 1994-1997. The dependent variable of the production frontier is the number of customers, and the regressors are: distribution lines (Kmnet), number of employees in distribution (Empl), service area (Area), transformer capacity (Transf), and proportion of sales to residential customers (Struct). Just as with the first model, we will treat this model as a cross-section one:

Model 4 (Energy Distribution – South America): the dependent variable is Lncustomers					
Variable	**OLS**	**MOLS-HN**	**MOLS-EXP**	**ML-HN**	**ML-EXP**
Lnempl	0.236 (4.23)	0.236	0.236	0.236 (4.23)	0.236 (4.23)
Lnkmnet	0.304 (9.66)	0.304	0.304	0.304 (9.66)	0.304 (9.66)
Lntransf	0.316 (9.64)	0.316	0.316	0.316 (9.64)	0.316 (9.64)
Lnstruct	-0.149 (-1.67)	-0.149	-0.149	-0.149 (-1.67)	-0.149 (-1.67)
Lnarea	0.009 (0.51)	0.009	0.009	0.009 (0.51)	0.009 (0.51)
Time	0.033 (1.11)	0.033	0.033	0.033 (1.11)	0.033 (1.11)
Constant	4.687 (10.07)	4.687	4.687	4.687 (10.07)	4.687 (10.07)
Gamma	- -	0	- -	0	- -
% of noise	- -	100%	100%	100%	100%

The ratio of the estimated coefficient to the standard error is presented between parentheses.
Ln stands for natural logarithm

Contrary to the first model, we have here a frontier with pure random noise where all the firms are considered 100% efficient. This is so because the skewness of the error term has the wrong sign:[16]

Figure 4



Estimated error term

In this case, the Jarque-Bera statistic is 3.38, and so we cannot reject the null that the OLS residuals have a normal distribution, at the usual level of confidence. As indicated in

[16] Therefore ML and MOLS estimates are the same.

section II, if the shape of the estimated composed error term looks like a normal distribution, this could be interpreted as a sign of no inefficiency in the model.

Consistency of results

We can observe that, in all cases, the Exponential distribution identifies a larger proportion of noise in the model than the H-N (as proposition 2 established). Moreover, our results corroborate what we have mentioned in section II about the Exponential distribution having a tendency to identify a larger number of efficient firms. The next table summarizes our findings:

Table 1

	MOLS-MN	**MOLS-EXP**	**ML-MN**	**ML-EXP**
Model 1 (Gas – Argentina)				
Mean	0.914	0.947	0.908	0.924
Deviation	0.050	0.031	0.075	0.077
Model 2 (Water – Asia and the Pacific Region)				
Mean	1.291	1.162	1.469	1.086
Deviation	0.117	0.063	0.328	0.023
Model 3 (Water – England)				
Mean	1.108	1.062	1.155	1.047
Deviation	0.055	0.031	0.115	0.022

It must be clear that the interpretation of the efficiency measures in a production frontier is not similar to the interpretation in a cost frontier. In the former case, the measure is between zero and one, showing the fraction of the maximum attainable product that the firm is actually producing.[17] In contrast, the measure of a cost frontier is larger than one and represents the percentage of cost over the minimum attainable.[18]

In order to check whether the populations from which the samples were extracted have identical population medians, we performed a Kruskal-Wallis (non-parametric) test on the four medians of each model, and for MOLS and ML individually. Except the ML estimates of the first model, in the remaining eight cases we were able to reject the null that the populations have identical population medians. This fact would be pointing to a lack of consistency between the distributions and so we can conclude that the median efficiency is sensitive to the assumed distribution.

Table 2 contains the coefficients of Spearman's rank correlations (SP statistic) for the rankings obtained in the ML-HN and ML-EXP estimations.[19]

Table 2

	Model 1	**Model 2**	**Model 3**
SP statistic	0.959	0.993	0.987

All the correlations are positive, and significantly different from zero at a 1% confidence level. That is, the ranking of firms is not substantially affected by the assumed distribution.

Finally, table 3 shows the fraction of firms that both estimators (ML-HN and ML-EXP) simultaneously classified in the upper and lower quartile. It is worthwhile noting that if the fraction were purely random, it would be expected to be around 25%.

[17] A measure of 0.8 means that the firm is producing 80% of the maximum attainable.

[18] A measure of 1.20 means that the firm is producing with costs that exceed in 20% the efficient level.

[19] We did not do the same for the MOLS estimates because the ranking of firms in this fashion is not affected by the distribution assumed. This is because the size of the composed error term always increases with the size of the OLS residual.

Table 3

	Model 1	Model 2	Model 3
First quartile	0.90	0.92	0.86
Fourth quartile	0.80	0.85	1.00

These results imply that both distributions identify the same firms as the “best” and the “worst”.

All these results corroborate what Kumbhakar and Lovell (2000) mentioned. These authors argue that there is empirical evidence showing that the sample mean efficiency is apt to be sensitive to the distribution assigned to the one-sided error component u_i. Nevertheless, they also note that what is not so clear is whether a ranking of producers by their individual efficiency scores, or the composition of the top and bottom efficiency scores deciles, is sensitive to distributional assumptions. They provide an example based on the data used by Greene (1990) where the H-N distribution and the Exponential distribution arrive basically at the same results.

We also used the data in Greene (1990) and we detected that the ML estimator finds a higher percentage of noise in the H-N case (56%) than in the Exponential case (52%), though the Exponential has higher sample mean efficiency than the H-N. On the other hand, the results for the MOLS estimator corroborate the three propositions in Section II. This example shows that our propositions are only valid for the MOLS estimator.

There are many other empirical works that used both distributions and found that the Exponential distribution had higher mean efficiencies than the H-N distribution. Jaforullah and Devlin (1996) used ML and found that the mean efficiency was higher (0.911) in the exponential case than in the H-N case (0.867). They estimated a production function and so technical efficiency. Jaforullah (1996) also used ML (production function), for a set of companies of Bangladesh and the results are mixed, though in general the Exponential had higher mean efficiencies. Parikh and Ali (1995) used ML to estimate a cost function, finding that the average inefficiency was 11.5% for the H-N case and 10.5% for the Exponential case. Parikh and Shah (1996) used ML (production function) and found that the Exponential distribution had a higher mean (97.4) than the H-N distribution (95.6). Ruggiero and Vitaliano (1999) estimated a cost function with ML technique and assumed an Exponential distribution. However, the authors argued that the Exponential consistently estimated less inefficiency than the H-N distribution. Finally, Bhavani (1991) performed MOLS (production function) and found that the Exponential distribution always had higher means that the H-N distribution. He highlighted this result, but he did not mention any theoretical justification.

IV. Conclusions

In this paper we have presented and analyzed two of the most commonly used distributions for the inefficiency term in a stochastic frontier: the Half Normal and the Exponential distributions, and we have performed an empirical application using four databases already used in previous work.

Our results corroborate what Kumbhakar and Lovell (2000) already mentioned, since the mean efficiency has been sensitive to the distribution assumed, whereas neither the rankings nor the top and bottom quartiles of the efficiency rankings have been affected. Moreover, in all cases we checked the results in section II, where we showed that the Exponential distribution has a tendency to identify a larger number of efficient firms than the H-N distribution.

It is worth mentioning that our results are quite robust, because we estimated different kinds of frontiers (production and cost) in diverse sectors (gas, water and electricity).

The fact that the mean efficiency changes with the distributional assumption implies that it would not be possible to translate the efficiency scores for each firm one-for-one into X factors. A firm's efficiency score that remained in the 90-100% interval, could be in the 80-90% interval under another distributional assumption. If we are looking for a measure that is not influenced by subjective considerations, then it is not recommendable to use the individual efficiency scores for regulatory objectives. However, both distributions recognize the same firms as the “best” and the “worst”, and so the alternative of publishing the results in the media as a mechanism of punishment is totally viable (see OFWAT, 1998). Likewise, given that the rankings are quite similar, it would be possible to order the firms by their relative position and then discriminate X factors according to this. This option would be analogous to the discrimination of X factors made by OFWAT according to a number of bands obtained from an average function. Actually, identifying the rough ordering of efficiency levels by firms is usually more important for regulatory policy decisions than measuring the level of efficiency itself. Therefore, stochastic frontier studies remain a potentially useful tool for regulatory purposes.

REFERENCES

Aigner, D., Lovell, C. and Schmidt, P. (1977): Formulation and estimation of stochastic frontier production function models, Journal of Econometrics, Vol. 6.

Bhavani, T. (1991): Technical Efficiency in Indian Modern Small Scale Sector: An Application of Frontier Production Function, Indian Economic Review, Vol. 26.

CEER (2000): The relative Performance of Electricity Distribution Firms in South America, Mimeo CEER.

Estache A. and Rossi M. (1999): How Different is the Efficiency of Public and Private Water Companies in Asia?, Policy Research Working Paper Nº 2152, The World Bank, July 1999.

Farrell M. (1957): The Measurement of Productive Efficiency, Journal of the Royal Statistical Society, Series A, Part III, Vol. 120.

Grenne, W. (1993): A gamma-distributed stochastic frontier model, Journal of Econometrics, Vol. 46.

Jaforullah, M. (1996): Technical Efficiencies of Some Manufacturing Industries of Bangladesh: An Application of the Stochastic Frontier Production Function Approach, Bangladesh Development Studies, Vol. 24.

Jaforullah, M. and Devlin, N. (1996): Technical Efficiency in the New Zealand Diary Industry: A Frontier Production Function Approach, Journal of Policy Modeling, Vol. 18.

Jondrow, J., Lovell, C., Materov, I. and Schmidt, P. (1982): On the estimation of technical inefficiency in the stochastic frontier production function model, Journal of Econometrics, Vol. 19.

Kumbhakar, S. and Lovell, C. (2000): Stochastic frontier analysis (Cambridge University press).

Lovell, C. (1993): Production frontiers and productive efficiency, in The measurement of productive efficiency, Fried, Lovell and Schmidt (Oxford University Press, New York).

Meeusen, W. and Van den Broeck, J. (1977): Efficiency estimation from Cobb-Douglas production functions with composed error, International Economic Review, Vol 18, Nº 2, June.

OFWAT (1998): Assessing the Scope for Future Improvements in Water Company Efficiency: A Technical Paper, Office of Water Services, Birmingham, UK, June.

Olson, J., Schmidt, P. and Waldman, D. (1980): A Monte Carlo study of estimators of the stochastic frontier production function , Journal of Econometrics, Vol. 13.

Parikh, A. and Shah, K. (1996): Various Approaches to Measurement of Technical Efficiency in North-West Frontier Province of Pakistan, Journal of Applied Economics, Vol. 12.

Parikh, A. and Ali, F. (1995):Measurement of Economic Efficiency in Pakistani Agriculture, American Journal of Agricultural Economics, Vol. 77.

Ritter, C. and Simar, L. (1997): Pitfalls of Normal-Gamma stochastic frontier models, Journal of Productivity Analysis 8: 2 (May).

Rossi M. (2001): Technical Change and Efficiency Measures: The Post-Privatisation in the Gas distribution Sector in Argentina, Energy Economics, Vol. 23.

Ruggiero, J. and Vitaliano, D. (1999): Assessing the Efficiency of Public Schools Using Data Envelopment Analysis and Frontier Regression, Contemporary Economic Policy, Vol. 17.

Shleifer, A. (1985): A Theory of Yardstick Competition, Rand Journal of Economics, Vol. 16, 3, Autumn, 319-327.

Stevenson, R. (1980): Likelihood functions for generalized stochastic frontier estimations, Journal of Econometrics, Vol. 13.

Stewart, M. (1993): Modelling Water Cost 1992-93: further research into the impact of operating conditions on company cost, OFWAT Research paper number 2, December.

Waldman D. (1982): A Stationary Point for the Stochastic Frontier Likelihood, Journal of Econometrics, Vol. 28.

CEER Working Paper Series

To order any of these papers, or all of these, see instructions at the end of the list. A complete list of CEER Working Papers is displayed here and in our web site

---

WPS 1. Laffont, Jean Jacques: Translating Principles Into Practice in Regulation Theory (March 1999)

WPS 2. Stiglitz, Joseph: Promoting Competition in Telecommunications (March 1999)

WPS 3. Chisari, Omar, Antonio Estache, y Carlos Romero: Winners and Losers from Utility Privatization in Argentina: Lessons from a General Equilibrium Model (March 1999)

WPS 4. Rodríguez Pardina, Martín y Martín Rossi: Efficiency Measures and Regulation: An Illustration of the Gas Distribution Sector in Argentina (April 1999)

WPS 5. Rodriguez Pardina, Martín Rossi and Christian Ruzzier: Consistency Conditions: Efficiency Measures for the Electricity Distribution Sector in South America (June 1999)

WPS 6. Gordon Mackerron: Current Developments and Problems of Electricity Regulation in the European Union and the United Kingdom (November 1999)

WPS 7. Martín Rossi: Technical Change and Efficiency Measures: The Post-Privatisation in the Gas Distribution Sector in Argentina (March 2000)

WPS 8. Omar Chisari, Martín Rodriguez Pardina and Martín Rossi: The Cost of Capital in Regulated Firms: The Argentine Experience (May 2000)

WPS 9. Omar Chisari, Pedro Dal-Bó and Carlos Romero: High Tension Electricity Network Expansions in Argentina: Decision Mechanisms and Willingness-to-Pay Revelation (May 2000).

WPS 10. Daniel A. Benitez, Antonio Estache, D. Mark Kennet, And Christian A. Ruzzier. Potential Role of Economic Cost Models in the Regulation of Telecommunications in Developing Countries (August 2000).

WPS 11. Martín Rodríguez Pardina and Martín Rossi. Technical Change and Catching-up: The Electricity Distribution Sector in South America

WPS 12. Martín Rossi and Iván Canay. Measuring Inefficiency in Public Utilities: Does the Distribution Matter?

Centro de Estudios Económicos de la Regulación

Subscription Request CEER's Publications

Please, include my name in the mailing list:

a) Working Papers Series (...) paper (...) e-mail, format pdf

b) Serie de Textos de Discusión (...) paper (...) e-mail, format pdf

My name is:...

Profession:...

Adress:...

....

....

Signature

Mail to:
SECRETARIA CEER
Lima 717, 1° piso
C1053AAO Buenos Aires
Argentina

Fax, 54-11-43797588
E-mail: ceer@uade.edu.ar